REMOTE SENSING LETTERS, 2016
VOL. 7, NO. 4, 358–367
http://dx.doi.org/10.1080/2150704X.2016.1142678

Taylor & Francis
Taylor & Francis Group

Combining transductive and active learning to improve object-based classification of remote sensing images

Fabio N. Güttler [a,b], Dino Ienco[b], Pascal Poncelet[c] and Maguelonne Teisseire[b]

[a]ICube, Université de Strasbourg, Strasbourg, France; [b]UMR TETIS, Irstea, Montpellier, France; [c]LIRMM, Université de Montpellier, Montpellier, France

ABSTRACT

In this letter, we propose a new active transductive learning (ATL) framework for object-based classification of satellite images. The framework couples graph-based label propagation with active learning (AL) to exploit positive aspects of the two learning settings. The transductive approach considers both labelled and unlabelled image objects to perform its classification as they are all available at training time while the AL strategy smartly guides the construction of the training set employed by the learner. The proposed framework was tested in the context of a land cover classification task using RapidEye optical imagery. A reference land cover map was elaborated over the whole study area in order to get reliable information about the performance of the ATL framework. The experimental evaluation underlines that, with a reasonable amount of training data, our framework outperforms state of the art classification methods usually employed in the field of remote sensing.

ARTICLE HISTORY
Received 17 July 2015
Accepted 9 January 2016

1. Introduction

Data labels are usually difficult and expensive to obtain. Standard classification techniques heavily rely on the hypothesis that a big quantity of labelled examples (training set) is available to build predictive models. Considering the remote sensing domain, in particular the object-based image classification, the label acquisition constitutes a time and effort consuming task for the expert. The collection of such labels can affect negatively the image classification task from two points of view: the quantity of labelled data commonly needed by standard inductive classifier and the way the objects to label are chosen.

Classical supervised inductive classification approaches (i.e. Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), etc.) require many labelled data to train the model. Also, they assume that training and test data are not available at the same time since the model they have learnt needs to be enough general to classify new unseen examples available in a near future (Vapnik 1998). However, in the case of remote sensing image classification, training examples are limited and all the objects (training and test) are available at the same time.

A different classification setting is supplied by transductive learning (Joachims 1999), which belongs to the family of the semi-supervised approaches. Transductive learning tries

CONTACT Fabio N. Güttler guttler@unistra.fr

to propagate information from the labelled data to the unlabelled one leveraging the availability of training and test data at the same time. These kinds of techniques offer an effective approach to supply contextual classification of unlabelled examples by using a relatively small set of labelled examples. Many real-world applications can be modelled through a transductive setting. In particular, it has been also applied in the remote sensing domain (Sun et al. 2014) where labels are difficult to obtain and the classification decisions should not be made separately from learning the current data. Differently from inductive classification, transductive learning does not produce any reusable model.

The second issue regards the way labels are collected. Objects which are usually labelled almost randomly by an expert while choosing examples guided by the needs of the classifier can drastically improve classification performance (Demir, Minello, and Bruzzone 2014). This kind of technique is called active learning (AL), and it allows to involve expert interaction during the classifier construction. More in detail, the choice of the objects to label is guided by the learner needs; therefore, labels are obtained for the objects that can, potentially, improve the performance of the model. The objects are generally selected considering the classifier uncertainty over the set of possible examples to choose (Fu, Zhu, and Li 2013). In remote sensing applications, this approach is getting more and more attention (Demir, Minello, and Bruzzone 2014) due to the improvement it can supply in the task of image classification.

In this letter, we propose to couple transductive and AL in order to design a new active transductive learning (ATL) framework. We adapted a label propagation approach (Liu and Chang 2009) to object-based image classification and combined it with an effective AL strategy. The proposed methodology was experimented in the context of land cover object-based image classification.

2. Study area and data set description

2.1. *Study area*

Experiments were performed on the *Lower Aude Valley* site, located in the south of France. Spanning over a coastal wetland area of about 4842 ha, this site is part of the European network of protection areas called Natura 2000. Most of the site (56.3%) is composed of natural and semi-natural areas, specially salt-meadows, salt-marshes and coastal lagoons. The rest (43.7%) is mainly occupied by agricultural parcels (vineyards, cereal crops, orchards) and some small built-up areas (roads and houses).

2.2. *Data set preparation*

As input raster data we used a RapidEye multispectral image acquired in 24 June 2009 and available in the context of the Geoinformation for Sustainable Development (GEOSUD) project. The image was provided in level 3A, which means that radiometric, sensor and geometric corrections were performed. At this processing level, the RapidEye product has a pixel spacing of 5 m and contains five spectral bands (approximate centre in nm): blue (475), green (555), red (657), red-edge (710) and near-infrared (805).

Image segmentation was performed using the five spectral bands and considering only the area inside the *Lower Aude Valley* site. This step is of great importance as it provides a new and more meaningful representation of the image. Instead of the arbitrary pixel grid, segmentation aims to create spatially coherent objects based on spectral and spatial features of adjacent pixels over the image. The segmentation was performed using the multiresolution

segmentation algorithm (MSA) available at *eCognition Developer 8.8.1*. MSA is a bottom-up segmentation based on a pairwise region merging technique. It uses a homogeneity criterion (combination of spectral and shape criteria) to decide whether to merge or not neighbouring pixels or objects. In our case, we set the MSA user parameters as following: scale = 100, shape = 0.2, compactness = 0.5. It generated a set of 13,292 objects with a good compromise in terms of under and over segmentation (Troya-Galvis et al. 2015). For each object, we calculated the following attributes: mean value for the five inner spectral bands and for five additional spectral indices – normalized difference vegetation index (NDVI) (Rouse et al. 1974), red edge normalized difference vegetation index ($NDVI_{re}$) (Gitelson and Merzlyak 1994; Kross et al. 2015), red edge normalized difference water index ($NDWI_{re}$) (Klemenjak et al. 2012), red edge triangular vegetation index ($RTVI_{core}$) (Chen et al. 2010) and simple ratio (SR) (Jordan 1969).

In parallel, the RapidEye image was used to create a reference map over the whole study site (see Figure 1). This task was carried out through a manual land cover digitalization and interpretation process at the scale of 1:10,000. Field surveys and precise aerials photographs (0.5 m of spatial resolution) were employed to ensure the exactness of the land cover map. Two specific sets of land cover classes were used label the individual map units (polygons in our case). The first set is specific to natural and semi-natural areas (from now called natural data set) while the second set concerns artificial, cultivated and managed areas (from now called artificial data set).

Finally, the land cover information (from the reference map) was propagated into the set of objects (from the image segmentation). Only the objects fitting completely inside the polygons of the reference map received a land cover label (see Figure 1). In total, 3357 objects were labelled for the natural data set and 3637 for the artificial data set. As the experiment reproduces a real task of land cover image classification, the number of objects

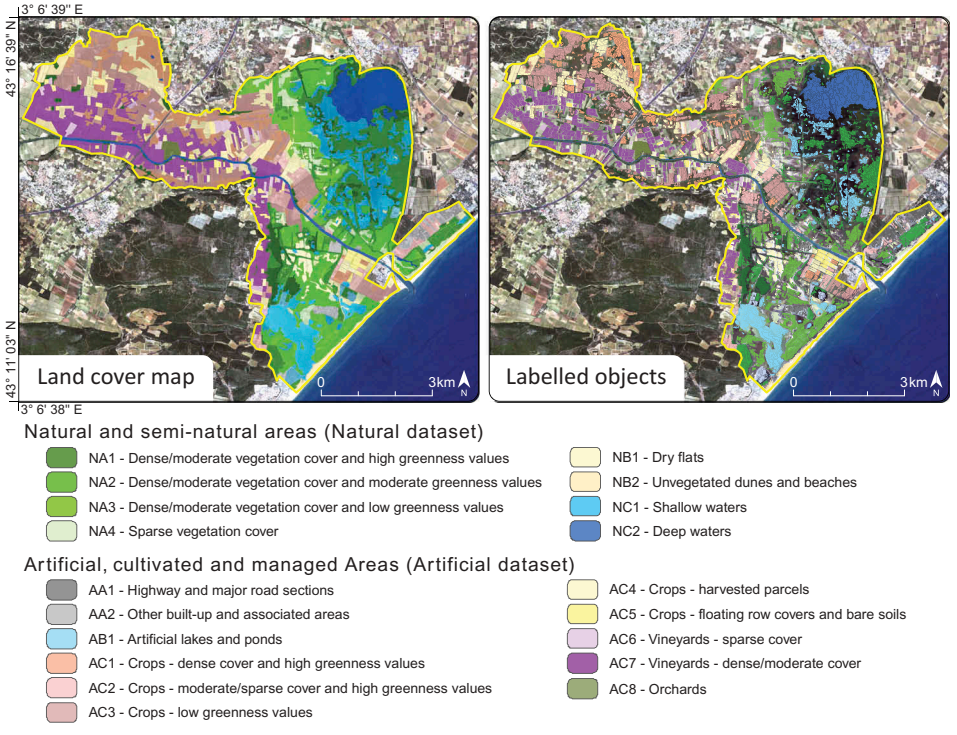

Figure 1. Land cover reference map for the Lower Aude Valley site (left side) and the spatial distribution of the labelled objects generated by the segmentation (right side).

per class is strongly unbalanced as one can notice in the following lists (the number of objects is indicated in brackets). Natural classes: NA1(264), NA2(760), NA3(1019), NA4(161), NB1(253), NB2(155), NC1(529) and NC2(216). Artificial classes: AA1(21), AA2(418), AB1(77), AC1(439), AC2(277), AC3(658), AC4(542), AC5(209), AC6(75), AC7(900) and AC8(21).

3. Methodology

In this section, we introduce the different components used to implement the ATL framework: (i) the transductive setting, (ii) the label propagation algorithm and (iii) the AL strategy.

3.1. *Transductive setting*

Given a set of objects $O = \{o_i\}_{i=1}^{N}$, let us denote with L the subset of labelled objects of O, and with U the subset of unlabelled ones. U can have any proportion w.r.t. L, but in many real cases U is much larger than L. Every object in L is associated to one class belonging to $C = \{C_j\}_{j=1}^{M}$, where M is the number of possible classes. We also denote with $\mathbf{Y}$ a $N \times M$ matrix such that the (i, j)th element of this matrix, $\mathbf{Y}_{ij}$, equals 1 if C_j is the label assigned to object o_i; $\mathbf{Y}_{ij} = 0$ otherwise. Without loss of generality, we can refer to L as training data and to U as test data.

The goal of a transductive learner (TL) is to make an inference 'from particular to particular', i.e. given the classifications of the instances in the training set L, it aims to predict the classifications of the instances in the test set U, rather than inducing a general rule that works out for classifying new unseen instances (Vapnik 1998). Transduction is naturally related to the class of case-based learning algorithms, whose most well-known algorithm is the k-nearest neighbour (k-NN) (Joachims 2003). Differently from standard supervised setting, in the transductive setting there is no separation between model training and testing phase. The classification of new unseen example is performed at the same time the model is learnt over L.

3.2. *Label propagation algorithm*

In order to perform transductive learning, we adapted the approach named robust multi-class graph transduction (RMGT) proposed by Liu and Chang (2009). From the best of our knowledge, it is the first time this kind of approach is employed in a remote sensing application, in particular to perform an object-based classification.

Essentially, RMGT implements a graph-based label propagation approach, which exploits a k-NN graph built over the entire data set to propagate the class information from the labeled to the unlabelled examples. The assumption behind this approach is that adjacent vertices are likely to have similar labels. For this reason, the label propagation procedure ensures that the classification function varies smoothly along the edges of the k-NN graph. In the following, we describe in detail the mathematical aspects of RMGT.

Let $G = \langle V, E, w \rangle$ be an undirected graph whose vertex set is $V = O$, edge set is $E = \{(o_i, o_j) | o_i, o_j \in O \wedge \mathrm{sim}(o_i, o_j) > 0\}$, and edge weighting is $w = \mathrm{sim}(o_i, o_j)$, where sim $(\cdot,\cdot)$ is defined as $\frac{1}{1+\mathrm{dist}(\cdot,\cdot)}$ and dist$(\cdot,\cdot)$ is the Euclidean distance between the feature vectors of two objects.

Input: A collection of objects O, with labelled objects L and unlabelled objects U (with $D = L \cup U$ and $L \cap U = \emptyset$);

Output: A classification over C for the objects in U.

(1) Build the similarity graph G for the object set O.
(2) Extract the k-nearest neighbour graph G_k from G. /* *Section 3.2* */
(3) Build the matrix $\mathbf{W}$ from G_k, which represents the symmetry-favoured k-NN graph. /* *Section 3.2* */
(4) Compute the normalized Laplacian of $\mathbf{W}$. /* *Section 3.2* */
(5) Compute the *RMGT* solution $\mathbf{F}$. /* *Equation 2* */
(6) Assign object $o_i \in U$ to the class C_j^* that maximizes the class likelihood, $j^* = \arg\max_j \mathbf{F}_{ij}$.

Algorithm 1: Object-Based Transductive Classification

Given a positive integer k, consider the k-NN graph $G_k = \langle V, E_k, w \rangle$ derived from G and such that $E = \{(d_i, d_j) | d_j \in N_i\}$, where N_i denotes the set of d_i's k-nearest neighbours. A weighted sparse matrix is obtained as $\mathbf{W} = \mathbf{A} + \mathbf{A}^{\mathrm{T}}$, where $\mathbf{A}$ is the weighted adjacency matrix of G_k and $\mathbf{A}^{\mathrm{T}}$ is the transpose of $\mathbf{A}$; the matrix $\mathbf{W}$ represents a *symmetry-favoured* k-NN graph (Liu and Chang 2009). Moreover, let $\mathbf{P} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ be the normalized Laplacian of $\mathbf{W}$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{D} = \mathrm{diag}(\mathbf{W}\mathbf{1}_N)$, where $\mathbf{1}_N$ is a vector of 1 s of size N. Without loss of generality, we can rewrite $\mathbf{P}$ and $\mathbf{W}$ as subdivided into four and two submatrices, respectively:

$$\mathbf{P} = \begin{bmatrix} \mathbf{\Delta}_{LL} & \mathbf{\Delta}_{LU} \\ \mathbf{\Delta}_{UL} & \mathbf{\Delta}_{UU} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_L \\ \mathbf{Y}_U \end{bmatrix} \tag{1}$$

where $\mathbf{\Delta}_{LL}$ and $\mathbf{Y}_L$ are the submatrices of $\mathbf{P}$ and $\mathbf{Y}$, respectively, corresponding to the labelled objects. More in detail, $\mathbf{\Delta}_{LL}$ contains the similarity between each pair of examples belong to the set of labelled objects while $\mathbf{\Delta}_{UL}$ contains all the distances between the unlabelled objects and the labelled one. The same analogy applies for all the other submatrices. The RMGT learning algorithm finally yields a matrix $\mathbf{F} \in R^{N \times M}$ defined as

$$\mathbf{F} = -\mathbf{\Delta}_{UU}^{-1}\mathbf{\Delta}_{UL}\mathbf{Y}_L + \frac{\mathbf{\Delta}_{UU}^{-1}\mathbf{1}_{|U|}}{\mathbf{1}_{|U|}^{\mathrm{T}}\mathbf{\Delta}_{UU}^{-1}\mathbf{1}_{|U|}}\left(N\omega - \mathbf{1}_{|L|}^{\mathrm{T}}\mathrm{Y}_L + \mathbf{1}_{|U|}^{\mathrm{T}}\mathbf{\Delta}_{UU}^{-1}\mathbf{\Delta}_{UL}\mathbf{Y}_L\right), \tag{2}$$

where $\omega \in R^M$ is the class probability distribution that is assumed uniform.

The learning schema used by RMGT employs spectral properties of the k-NN graph to spread the labelled information over the set of test instances. Specifically, the label propagation process is modelled as a constrained convex optimization problem where the labelled objects are employed to constrain and guide the final classification. Equation 2 represents the closed form solution of the propagation process. This equation shows how labelled (L) and unlabelled (U) examples are combined to implement the main assumption that adjacent vertices are likely to have similar labels. After the propagation step, every unlabelled object o_i is associated to a vector (i.e. the ith row of $\mathbf{F}$) representing the likelihood of the object o_i for each of the classes; therefore, o_i is assigned to the class that maximizes the likelihood.

Algorithm 1 sketches the main steps of the approach. Initially, the similarity matrix between all the objects is computed (Line 1). The graph-based label propagation process requires the construction of the k-NN graph (Line 2) and its symmetry-favoured transformation (Line 3). After that, the algorithm computes the normalized Laplacian of the matrix (Line 4) and the RMGT algorithm is applied on such data matrix. Line 6 describes the decision rule we adopted to perform the classification.

3.3. *Active transductive learning*

AL is getting more attention in remote sensing image classification as it helps to deal with the time and effort consuming task of collecting a good quality training set to build a classification model (Demir and Bruzzone 2015; Demir, Minello, and Bruzzone 2014). The general AL loop (Fu, Zhu, and Li 2013) involves the interaction between the classifier and the expert. Firstly a budget is defined, it represents the percentage of examples the experts is willing to label. Then the AL loop starts. At each iteration, the procedure ranks the set of unlabelled examples in order to promote in the rank the more relevant examples to label. Each example with is ranked considering its importance to the current learnt classifier. Once the rank is produced, the procedure chooses the top objects (one or more) and asks to the expert their true labels. The selected objects are added to the current training set and the classifier is updated. The AL loop stops when the budget is exhausted.

In this work, we adopted the *margin-based* strategy to score examples. This heuristic is chosen as it usually obtains better results than other uncertainty sampling methods (Fu, Zhu, and Li 2013). This strategy considers the probability distribution of a classifier c on the example x over the possible set of classes C. It is prone to select instances with minimum margin between posterior probabilities of the two most likely class labels. More formally, it is defined as: $\text{margin}(\mathbf{x}) = P_c(\mathbf{x}|C_{\text{first}}) - P_c(\mathbf{x}|C_{\text{second}})$, where given the classifier c, $P_c(\mathbf{x}|C_i)$ is the probability of the classifier c to predict the class label C_i for the example $\mathbf{x}$, C_{first} is the most probable class for the example $\mathbf{x}$ and C_{second} is the second most probable class for the classifier c. Values of margin($\mathbf{x}$) close to 0 indicate big uncertainty on $\mathbf{x}$ while values close to 1 underline reliable confidence in the prediction. In the AL step, first the unlabelled instances are ranked in ascending order w.r.t. their margin value, then the top n examples are supplied to the expert and their true label is obtained. We fixed the number of examples at each loop equals to 20.

Considering our framework, we coupled the *margin-based* heuristic with Algorithm 1. Given an object to classify o_i, we employ the likelihood vector ($\mathbf{F}_i$) as posterior probability distribution to implement the *margin-based* strategy.

3.4. *Experimental setting*

We compared our proposal with respect to state of the art classification approaches. As competitors we used the RF classifier, the SVM and the NB approach. For SVM, we evaluated both radial basis function and polynomial kernel and, at the end, we chose polynomial kernel with exponent value equals to 8 as it supplied the best results. We coupled each of the competitors with the same AL strategy that we employed in our ATL framework (RF + AL, SVM + AL, NB + AL).

This was done in order to fairly compare our proposal with the competitors. We also investigated the benefit supplied by the AL step by comparing the performances of our ATL against the base TL. For all the competitors we used the *Weka*[1] implementation. For the RMGT a k value equals to 15 was used for building the k-NN graph. We varied the training percentage (budget) between 2% to 40% in steps of 2%. This percentage indicates the proportion of the data set that was randomly sampled to create the training set. We evaluated the classification performance using the F-measure (Gómez-Chova et al. 2011). We used F-measure instead of general accuracy due to its ability to better describe classifier performance on unbalanced data sets. We randomly initialized each classifier with an object per

class before of starting the AL process. For each pair 'classifier and training percentage' we reported the average results over 30 runs.

4. Results and discussion

Classification performances over the two data sets (Natural and Artificial) of the *Lower Aude Valley* site are synthesized on Figure 2. As a general trend, for both data sets, we can observe that the classification performance usually improves by increasing the number of objects available in the training set.

Considering the Natural data set, we can notice that ATL outperformed all the other algorithms for any training percentage. Among the competitors, RF clearly outperformed SVM and NB. The *F*-measure curves obtained for the ATL and RF approaches are similarly shaped but separated by a regular shift of almost 3 points. The NB *F*-measure curve has also the same general shape but the shift is around 10 points w.r.t. ATL. Conversely, the SVM curve presented a distinct shape as it becomes stable much earlier than the other approaches. With training percentages bigger than 12% only marginal improvements are obtained for SVM.

Over the artificial data set, we can observe that all the algorithms, excepted SVM, presented lower performances, for small training percentages, in comparison with the first data set. However, when the budget increases and reaches a reasonable percentage (equal or bigger than 20%) the general trend changes, ATL continues to improve its performance outperforming SVM. The maximum gap between ATL and SVM is around 7 points, achieved for a budget of 40%.

The two plots (b and d) of Figure 2 report the performance of our approach with and without AL. For training percentages bigger than 6% ATL clearly outperforms the base TL

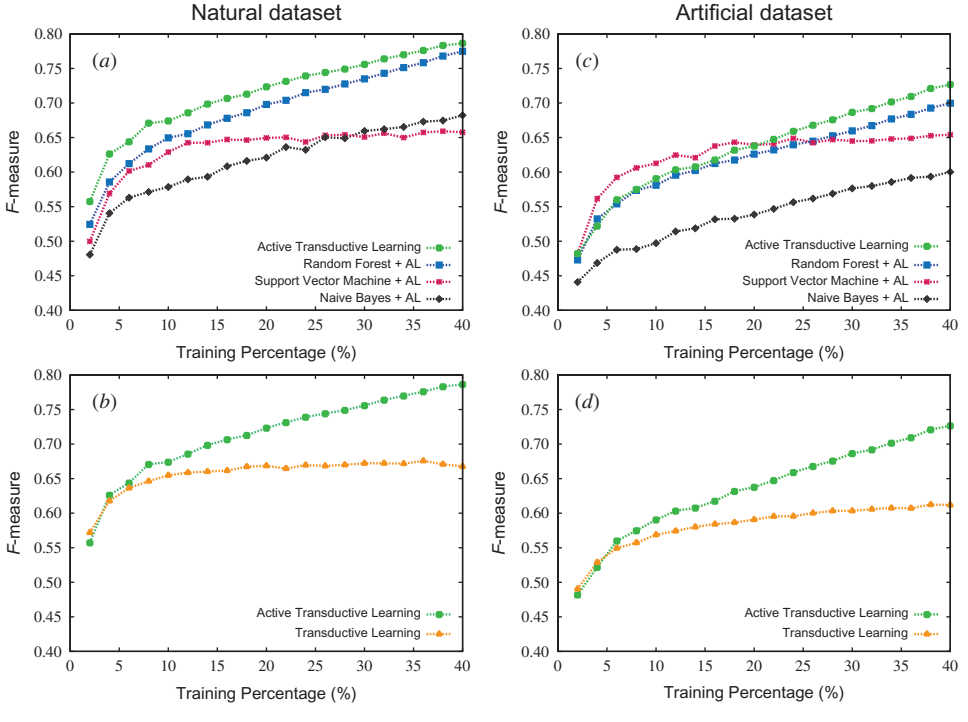

Figure 2. Classification results varying the percentage of training examples between 2% to 20% for: the Natural (left side) and the Artificial (right side) data sets. (a) All classification methods coupled with AL over the Natural data set. (b) Our framework with and without AL over the Natural data set. (c) All classification methods coupled with AL over the *Artificial* data set. (d) Our framework with and without AL over the *Artificial* data set.

underlining that building a training set guided by the classifier needs positively influences the final performance. Regarding smaller training percentages, we notice that the difference is very small. This fact points out that the benefit of AL only becomes evident when the training set size exceeds a minimum threshold, in our case between 6% and 8%.

In a more general evaluation, our experiments showed that the ATL framework outperforms all the state of the art methods in both data sets. However, considering the artificial data set with training percentages smaller than 20%, the SVM + AL obtained results that differ of 2 or 3 points from the ATL. The differences in the performances can be mainly explained by the nature of the two classifiers.

For both data sets, SVM + AL quickly reaches its maximum, and then it remains stable. In the first iterations, the AL strategy successfully selects the optimal support vectors allowing to define the best classification hyperplane. Once those instances are included in the training set, no more examples can improve the performance as they will not modify the classification hyperplane. Instead, considering our proposal, the AL step continuously selects useful examples that are employed by the transductive learner to improve the classification performance.

4.1. *Per class analysis of ATL performance*

In order to obtain a more accurate understanding of the ATL performances, we analysed the *F*-measure results per class for both data sets. This analysis is also useful to highlight some specificities of the data sets. Figure 3 shows per class performance obtained using a training percentage of 20% for the natural (a) and the artificial (b) data sets.

Considering the eight classes of the natural data set, three of them presented very good performances with the *F*-measure values ranging from 0.80 to 0.82 (sandy areas, shallow and deep waters). Four classes (mainly related to natural vegetation areas) showed intermediate performances ranging from 0.68 and 0.76 while only one class (NA4) presented poor results with less than 0.09. This class represents areas with sparse vegetation cover and is the smallest class of this data set (161 objects). The omission errors are dramatic here as most of the NA4 objects were classified as NA3 or NB1, which corresponds to low greenness vegetation areas and dry flats, respectively. These confusions are not really surprising as the sparse vegetation cover combines small vegetation patches surrounded by bare areas. Here, we should consider the influence of over-segmented areas: instead of creating large objects including 'patches and surrounding areas' the segmentation mostly isolated such small patches into separated objects.

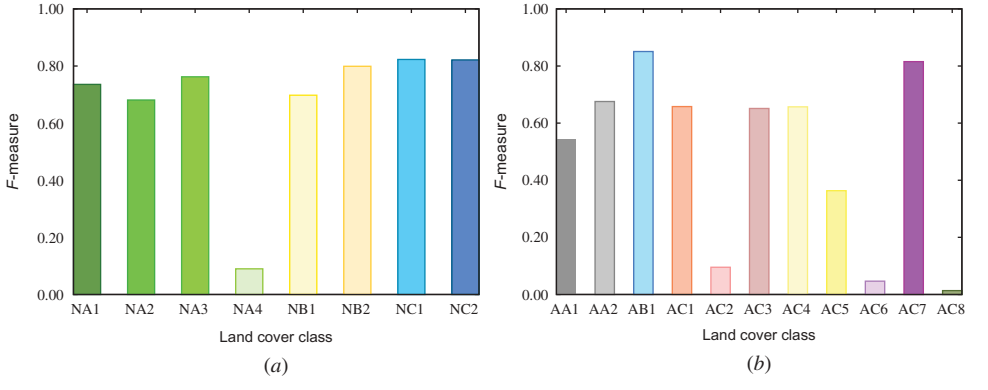

Figure 3. *F*-measure values for each class obtained with the ATL framework using a training percentage of 20%. (a) Natural data set classes. (b) Artificial data set classes.

Analysing the 11 classes of the artificial data set, we observe that two of them presented very good *F*-measure results (0.81 for vineyards with dense/moderate cover and 0.85 to artificial lakes and ponds). Five classes (concerning different crops and built-up areas) showed intermediate performances ranging from 0.54 to 0.68 and the four remaining classes presented poor results ranging from 0.01 to 0.36. All these four classes (AC2, AC5, AC6 and AC8) are characterized by extreme omission errors. The worst of them (AC8) possesses only 21 objects and corresponds to an orchard area. Most of their objects were classified as crops fields with vegetation (AC1 or AC3) what is somehow logical. During these periods of the year (early summer), the orchard presented a dense cover with a spectral response very near to some cereal crops. In fact, this class was created in the reference map mostly based in context and field survey information instead of the spectral information. This is also true for the AC6 class where vineyards with sparse cover were classified sometimes as vineyards with dense cover and sometimes as harvested parcels. For the AC2 class (crops with moderate/sparse cover and high greenness values) the omission errors are clearly related to the way the reference map was constructed. Usually, each AC2 polygon in the reference map groups a set of very small and adjacent agricultural parcels separated by narrow, but very greenness, borders. However this level of generalization was not attained in the image segmentation step. Instead, it generated many small objects isolating such reduced strips of greenness vegetation from the core of the parcels. As a consequence, these objects were mainly classified as AC3 (core of the parcels) or AC1 (borders). Finally, the AC5 class, which represents crop fields with very high reflectance values (floating row covers and bare soils), was mainly classified as built-up and associated areas (AA2). This confusion is frequent in optical remote sensing as some of the materials used in civil construction can present spectral signatures close to those of bare soils and floating row covers.

5. Conclusion

In this letter, we presented a new ATL framework that efficiently deals with object-based image classification. While standard classification techniques employ inductive learning, ATL exploits training and test sets together actively choosing new examples that can improve the final classification. The proposed approach was experimented over the *Lower Aude Valley* study area using a RapidEye satellite image. A reference land cover map was elaborated over the whole study area and permitted a detailed assessment of the performance of the ATL framework. The proposed approach outperformed the competitors over two data sets considering a reasonable amount of training data. As future work we would investigate more sophisticated AL techniques leveraging spatial autocorrelation. To overpass some limitations related to the generalization of the experimental findings we plan to extend our work over other study sites.

Note

1. http://www.cs.waikato.ac.nz/ml/weka/

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the French National Research Agency in the framework of the program 'Investissements d'Avenir' [GEOSUD project, ANR-10-EQPX-20].

ORCID

Fabio N. Güttler http://orcid.org/0000-0003-2285-4122

References

Chen, P.-F., N. Tremblay, J.-H. Wang, P. Vigneault, W.-J. Huang, and B.-G. Li. 2010. "New Index for Crop Canopy Fresh Biomass Estimation." *Spectroscopy and Spectral Analysis* 30 (2): 512–517.

Demir, B., and L. Bruzzone. 2015. "A Novel Active Learning Method in Relevance Feedback for Content-Based Remote Sensing Image Retrieval." *IEEE Transactions on Geoscience and Remote Sensing* 53 (5): 2323–2334. doi:10.1109/TGRS.2014.2358804.

Demir, B., L. Minello, and L. Bruzzone. 2014. "An Effective Strategy to Reduce the Labeling Cost in the Definition of Training Sets by Active Learning." *IEEE Geoscience and Remote Sensing Letters* 11 (1): 79–83. doi:10.1109/LGRS.2013.2246539.

Fu, Y., X. Zhu, and B. Li. 2013. "A Survey on Instance Selection for Active Learning." *Knowledge and Information Systems* 35 (2): 249–283. doi:10.1007/s10115-012-0507-8.

Gitelson, A., and M. N. Merzlyak. 1994. "Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation." *Journal of Plant Physiology* 143 (3): 286–292. doi:10.1016/S0176-1617(11)81633-0.

Gómez-Chova, L., J. Muñoz-Marí, V. Laparra, J. Malo-López, and G. Camps-Valls. 2011. "A Review of Kernel Methods in Remote Sensing Data Analysis." In *Optical Remote Sensing*, 171–206. Berlin: Springer. doi:10.1007/978-3-642-14212-3_10.

Joachims, T. 1999. "Transductive Inference for Text Classification using Support Vector Machines." In *Proceedings of the Sixteenth International Conference on Machine Learning (ICML-1999)*, Bled, June 27–30, 200–209. Morgan Kaufmann Publishers Inc.

Joachims, T. 2003. "Transductive Learning via Spectral Graph Partitioning." In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, DC, August 21–24, 290–297. Palo Alto, CA: AAAI Press.

Jordan, C. F. 1969. "Derivation of Leaf-Area Index from Quality of Light on the Forest Floor." *Ecology* 50 (4): 663–666. doi:10.2307/1936256.

Klemenjak, S., B. Waske, S. Valero, and J. Chanussot. 2012. "Unsupervised River Detection in Rapideye Data." In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 6860–6863. doi:10.1109/IGARSS.2012.6352587.

Kross, A., H. McNairn, D. Lapen, M. Sunohara, and C. Champagne. 2015. "Assessment of Rapideye Vegetation Indices for Estimation of Leaf Area Index and Biomass in Corn and Soybean Crops." *International Journal of Applied Earth Observation and Geoinformation* 34: 235–248. doi:10.1016/j.jag.2014.08.002.

Liu, W., and S.-F. Chang. 2009. "Robust Multi-Class Transductive Learning with Graphs." In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 381–388. doi:10.1109/CVPR.2009.5206871.

Rouse, J. W., Jr., R. H. Haas, J. A. Schell, and D. W. Deering. 1974. "Monitoring Vegetation Systems in the Great Plains with ERTS." *NASA Special Publication* 351: 309.

Sun, Z., C. Wang, D. Li, and J. Li. 2014. "Semisupervised Classification for Hyperspectral Imagery with Transductive Multiple-Kernel Learning." *IEEE Geoscience and Remote Sensing Letters* 11 (11): 1991–1995. doi:10.1109/LGRS.2014.2316141.

Troya-Galvis, A., P. Gancarski, N. Passat, and L. Berti-Equille. 2015. "Unsupervised Quantification of Under- and Over-Segmentation for Object-Based Remote Sensing Image Analysis." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (5): 1936–1945. doi:10.1109/JSTARS.2015.2424457.

Vapnik, V. 1998. *Statistical Learning Theory*. New York: Wiley.